Unless specifically mentioned, please do all the work in R.

**Problem 1.**

Investigate the **cheddar** data set in the faraway package. In other words, read it in, check the information in it, check its dimension, and look at a few rows.

1. State the mean for the *taste* variable, the maximum value of the *Acetic* variable, the median for the *H2S* variable, and the 1st quartile (Q1) for the *Lactic* variable.
2. With *taste* as the response, estimate the linear model that has *Acetic*, *H2S*, and *Lactic* as predictors.
3. Create the model matrix for the three predictors (including the intercept). Call it . State its dimension and rank.
4. Create the projection matrix for , call it PX (do not put its contents in your answer; it is very large). Calculate and report the trace of PX.
5. Project the response (*taste*) onto the column space (manifold) of . Check the result; it should be a vector of dimension 30. What does this vector represent? There’s another way we could have obtained this vector. How?
6. State both the R-squared and the RSS for this model.
7. What would you expect the cross product of the fitted value vector and the residual vector to be? Calculate it. Were you correct?
8. What are the degrees of freedom of the model?
9. Estimate the variance of the errors of the model.
10. Which of the estimates of the three predictors is the most variable? (Check its standard error). Which one is most consistent?
11. Looking at the standard errors, t-values, and p-values of the three predictors, for which one is the true parameter (coefficient) most likely equal to zero?

**Problem 2.**

1. Letting *taste* be the response again, estimate the linear model that has no predictors (intercept only). Compare to your answer for problem 1a).
2. State the RSS for this model.
3. Look at the summary for this model and note that no is reported. Why not? What would be the for this model?
4. What is the reduction in RSS going from the model with no predictors to the model with 3 predictors?
5. Would you say that it makes sense to include the three predictors into the model?
6. What is the difference between an error and a residual?